COLON SPECIFIC GENES AND PROTEINS

This invention relates to newly identified polynucleotides, polypeptides encoded by such polynucleotides, and the use of such polynucleotides and polypeptides for detecting disorders of the colon, particularly the presence of colon cancer and colon cancer metastases. The present invention further relates to inhibiting the production and function of the polypeptides of the present invention. The thirteen colon specific genes of the present invention are sometimes hereinafter referred to as "CSG1", "CSG2" etc.

The gastrointestinal tract is the most common site of both newly diagnosed cancers and fatal cancers occurring each year in the USA, figures are somewhat higher for men than for women. The incidence of colon cancer in the USA is increasing, while that of gastric cancer is decreasing, cancer of the small intestine is rare. The incidence of gastrointestinal cancers varies geographically. Gastric cancer is common in Japan and uncommon in the United States, whereas colon cancer is uncommon in Japan and common in the USA. An environmental etiologic factor is strongly suggested by the statistical data showing that people who move to a high-risk area assume the high risk. Some of the suggested etiologic factors for gastric cancer include aflatoxin, a

carcinogen formed by *aspergillus flavus* and present in contaminated food, smoked fish, alcohol, and Vitamin A and magnesium deficiencies. A diet high in fat and low in bulk, and, possibly, degradation products of sterol metabolism may be the etiologic factors for colon cancer. Certain disorders may predispose to cancer, for example, pernicious anemia to gastric cancer, untreated non-tropical sprue and immune defects to lymphoma and carcinoma, and ulcerative and granulomatous colitis, isolated polyps, and inherited familial polyposis to carcinoma of the colon.

The most common tumor of the colon is adenomatous polyp. Primary lymphoma is rare in the colon and most common in the small intestine.

Adenomatous polyps are the most common benign gastrointestinal tumors. They occur throughout the GI tract, most commonly in the colon and stomach, and are found more frequently in males than in females. They may be single, or more commonly, multiple, and sessile or pedunculated. They may be inherited, as in familial polyposis and Gardener's syndrome, which primarily involves the colon. Development of colon cancer is common in familial polyposis. Polyps often cause bleeding, which may occult or gross, but rarely cause pain unless complications ensue. Papillary adenoma, a less common form found only in the colon, may also cause electrolyte loss and mucoid discharge.

A malignant tumor includes a carcinoma of the colon which may be infiltrating or exophytic and occurs most commonly in the rectosigmoid. Because the content of the ascending colon is liquid, a carcinoma in this area usually does not cause obstruction, but the patient tends to be to present late in the course of the disease with anemia, abdominal pain, or an abdominal mass or a palpable mass.

The prognosis with colonic tumors depends on the degree of bowel wall invasion and on the presence of regional lymph node involvement and distant metastases. The prognosis with

carcinoma of the rectum and descending colon is quite unexpectedly good. Cure rates of 80 to 90% are possible with early resection before nodal invasion develops. For this reason, great care must be taken to exclude this disease when unexplained anemia, occult gastrointestinal bleeding, or change in bowel habits develop in a previously healthy patient. Complete removal of the lesion before it spreads to the lymph nodes provides the best chance of survival for a patient with cancer of the colon. Detection in an asymptotic patient by occult-bleeding, blood screening results in the highest five year survival.

Clinically suspected malignant lesions can usually be detected radiologically. Polyps less than 1 cm can easily be missed, especially in the upper sigmoid and in the presence of diverticulosis. Clinically suspected and radiologically detected lesions in the esophagus, stomach or colon can be confirmed by fiber optic endoscopy combined with histologic tissue diagnosis made by directed biopsy and brush sitology. Colonoscopy is another method utilized to detect colon diseases. Benign and malignant polyps not visualized by X-ray are often detected on colonoscopy. In addition, patients with one lesion on X-ray often have additional lesions detected on colonoscopy. Sigmoidoscope examination, however, only detects about 50% of colonic tumors.

The above methods of detecting colon cancer have drawbacks, for example, small colonic tumors may be missed by all of the above-described methods. The importance of detecting colon cancer is also extremely important to prevent metastases.

In accordance with an aspect of the present invention, there are provided nucleic acid probes comprising nucleic acid molecules of sufficient length to specifically hybridize to the RNA transcribed from the human colon specific genes of the present invention or to DNA corresponding to such RNA.

In accordance with another aspect of the present invention there is provided a method of and products for diagnosing colon cancer metastases by detecting the presence of RNA transcribed from the human colon specific genes of the present invention or DNA corresponding to such RNA in a sample derived from a host.

In accordance with yet another aspect of the present invention, there is provided a method of and products for diagnosing colon cancer metastases by detecting an altered level of a polypeptide corresponding to the colon specific genes of the present invention in a sample derived from a host, whereby an elevated level of the polypeptide indicates a colon cancer diagnosis.

In accordance with another aspect of the present invention, there are provided isolated polynucleotides encoding human colon specific polypeptides, including mRNAs, DNAs, cDNAs, genomic DNAs, as well as antisense analogs and biologically active and diagnostically or therapeutically useful fragments thereof.

In accordance with still another aspect of the present invention there are provided human colon specific genes which include polynucleotides as set forth in the sequence listing.

In accordance with a further aspect of the present invention, there are provided novel polypeptides encoded by the polynucleotides, as well as biologically active and diagnostically or therapeutically useful fragments, analogs and derivatives thereof.

In accordance with yet a further aspect of the present invention, there is provided a process for producing such polypeptides by recombinant techniques comprising culturing recombinant prokaryotic and/or eukaryotic host cells, containing a polynucleotide of the present invention, under conditions promoting expression of said proteins and subsequent recovery of said proteins.

In accordance with yet a further aspect of the present invention, there are provided antibodies specific to such polypeptides.

In accordance with another aspect of the present invention, there are provided processes for using one or more of the polypeptides of the present invention to treat colon cancer and for using the polypeptides to screen for compounds which interact with the polypeptides, for example, compounds which inhibit or activate the polypeptides of the present invention.

In accordance with yet another aspect of the present invention, there are provided compounds which inhibit activation of one or more of the polypeptides of the present invention which may be used to therapeutically, for example, in the treatment of colon cancer.

In accordance with yet a further aspect of the present invention, there are provided processes for utilizing such polypeptides, or polynucleotides encoding such polypeptides, for in vitro purposes related to scientific research, synthesis of DNA and manufacture of DNA vectors.

These and other aspects of the present invention should be apparent to those skilled in the art from the teachings herein.

The following drawings are illustrative of embodiments of the invention and are not meant to limit the scope of the invention as encompassed by the claims.

Figure 1 is a partial cDNA sequence and the corresponding deduced amino acid sequence of a colon specific gene of the present invention.

Figure 2 is a partial cDNA sequence and the corresponding deduced amino acid sequence of a colon specific gene of the present invention.

Figure 3 is a partial cDNA sequence of a colon specific gene of the present invention.

Figure 4 is a partial cDNA sequence and the corresponding deduced amino acid sequence of a colon specific gene of the present invention.

Figure 5 is a partial cDNA sequence and the corresponding deduced amino acid sequence of a colon specific gene of the present invention.

Figure 6 is a partial cDNA sequence and the corresponding deduced amino acid sequence of a colon specific gene of the present invention.

Figure 7 is a partial cDNA sequence a colon specific gene of the present invention.

Figure 8 is a full length cDNA sequence and the corresponding deduced amino acid sequence of a colon specific gene of the present invention.

Figure 9 is a full length cDNA sequence and corresponding deduced amino acid sequence of a colon specific gene of the present invention.

Figure 10 is a partial cDNA sequence and corresponding deduced amino acid sequence of a colon specific gene of the present invention.

Figure 11 is a partial cDNA sequence and the corresponding deduced amino acid sequence of a colon specific gene of the present invention.

Figure 12 is a partial cDNA sequence of a colon specific gene of the present invention.

Figure 13 is a partial cDNA sequence of a colon specific gene of the present invention.

The term "colon specific gene" means that such gene is primarily expressed in tissues derived from the colon, and such genes may be expressed in cells derived from tissues other than from the colon. However, the expression of such genes is significantly higher in tissues derived from the colon than from non-colon tissues.

In accordance with one aspect of the present invention there is provided a polynucleotide which encodes one of the

mature polypeptides having the deduced amino acid sequence of Figure 8 or 9 and fragments, analogues and derivatives thereof.

In accordance with a further aspect of the present invention there is provided a polynucleotide which encodes the same mature polypeptide as a human gene having a coding portion which contains a polynucleotide which is at least 90% identical (preferably at least 95% identical and most preferably at least 97% or 100% identical) to one of the polynucleotides of Figures 1-7 or 9-13, as well as fragments thereof.

In accordance with still another aspect of the present invention there is provided a polynucleotide which encodes for the same mature polypeptide as a human gene whose coding portion includes a polynucleotide which is at least 90% identical to (preferably at least 95% identical to and most preferably at least 97% or 100% identical) to one of the polynucleotides included in ATCC Deposit No. 97,102 deposited March 20, 1995.

In accordance with yet another aspect of the present invention, there is provided a polynucleotide probe which hybridizes to mRNA (or the corresponding cDNA) which is transcribed from the coding portion of a human gene which coding portion includes a DNA sequence which is at least 90% identical to (preferably at least 95% identical to) and most preferably at least 97% or 100% identical) to one of the polynucleotide sequences of Figures 1-13.

The present invention further relates to a mature polypeptide encoded by a coding portion of a human gene which coding portion include a DNA sequence which is at lest 90% identical to (preferably at least 95% identical to and more preferably 97% or 100% identical to) one of the polynucleotides of Figures 1-7 or 10-13, as well as analogues, derivatives and fragments of such polypeptides.

The present invention also relates to one of the mature polypeptides of Figures ~~9 or 10~~ 8 or 9 and fragments, analogues and derivatives of such polypeptides.

The present invention further relates to the same mature polypeptide encoded by a human gene whose coding portion includes DNA which is at least 90% identical to (preferably at least 95% identical to and more preferably at least 97% or 100% identical to) one of the polynucleotides included in ATCC Deposit No. 97,102 deposited March 20, 1995.

In accordance with an aspect of the present invention, there are provided isolated nucleic acids (polynucleotides) which encode for the mature polypeptides having the deduced amino acid sequence of Figures 8 or 9 or fragments, analogues or derivatives thereof.

The polynucleotides of the present invention may be in the form of RNA or in the form of DNA, which DNA includes cDNA, genomic DNA, and synthetic DNA. The DNA may be double-stranded or single-stranded, and if single stranded may be the coding strand or non-coding (anti-sense) strand. The coding sequence which encodes the mature polypeptide may include DNA identical to Figures 1-13 or that of the deposited clone or may be a different coding sequence which coding sequence, as a result of the redundancy or degeneracy of the genetic code, encodes the same mature polypeptide as the coding sequence of a gene which coding sequence includes the DNA of Figures 1-13 or the deposited cDNA.

The polynucleotide which encodes a mature polypeptide of the present invention may include, but is not limited to: only the coding sequence for the mature polypeptide; the coding sequence for the mature polypeptide and additional coding sequence such as a leader or secretory sequence or a proprotein sequence; the coding sequence for the mature polypeptide (and optionally additional coding sequence) and non-coding sequence, such as introns or non-coding sequence

5' and/or 3' of the coding sequence for the mature polypeptide.

Thus, the term "polynucleotide encoding a polypeptide" encompasses a polynucleotide which includes only coding sequence for the polypeptide as well as a polynucleotide which includes additional coding and/or non-coding sequence.

The present invention further relates to variants of the hereinabove described polynucleotides which encode fragments, analogs and derivatives of a mature polypeptide of the present invention. The variant of the polynucleotide may be a naturally occurring allelic variant of the polynucleotide or a non-naturally occurring variant of the polynucleotide.

Thus, the present invention includes polynucleotides encoding the same mature polypeptide as hereinabove described as well as variants of such polynucleotides which variants encode a fragment, derivative or analog of a polypeptide of the invention. Such nucleotide variants include deletion variants, substitution variants and addition or insertion variants.

The polynucleotides of the invention may have a coding sequence which is a naturally occurring allelic variant of the human gene whose coding sequence includes DNA as shown in Figures 1-13 or of the coding sequence of the DNA in the deposited clone. As known in the art, an allelic variant is an alternate form of a polynucleotide sequence which may have a substitution, deletion or addition of one or more nucleotides, which does not substantially alter the function of the encoded polypeptide.

The present invention also includes polynucleotides, wherein the coding sequence for the mature polypeptide may be fused in the same reading frame to a polynucleotide sequence which aids in expression and secretion of a polypeptide from a host cell, for example, a leader sequence which functions as a secretory sequence for controlling transport of a polypeptide from the cell. The polypeptide having a leader

sequence is a preprotein and may have the leader sequence cleaved by the host cell to form the mature form of the polypeptide. The polynucleotides may also encode a proprotein which is the mature protein plus additional 5' amino acid residues. A mature protein having a prosequence is a proprotein and is an inactive form of the protein. Once the prosequence is cleaved an active mature protein remains.

Thus, for example, the polynucleotide of the present invention may encode a mature protein, or a protein having a prosequence or a protein having both a presequence and a presequence (leader sequence).

The polynucleotides of the present invention may also have the coding sequence fused in frame to a marker sequence which allows for purification of the polypeptide of the present invention. The marker sequence may be a hexa-histidine tag supplied by a pQE-9 vector to provide for purification of the mature polypeptide fused to the marker in the case of a bacterial host, or, for example, the marker sequence may be a hemagglutinin (HA) tag when a mammalian host, e.g. COS-7 cells, is used. The HA tag corresponds to an epitope derived from the influenza hemagglutinin protein (Wilson, I., et al., Cell, 37:767 (1984)).

The present invention further relates to polynucleotides which hybridize to the hereinabove-described polynucleotides if there is at least 70%, preferably at least 90%, and more preferably at least 95% identity between the sequences. The present invention particularly relates to polynucleotides which hybridize under stringent conditions to the hereinabove-described polynucleotides. As herein used, the term "stringent conditions" means hybridization will occur only if there is at least 95% and preferably at least 97% identity between the sequences. The polynucleotides which hybridize to the hereinabove described polynucleotides in a preferred embodiment encode polypeptides which either retain substantially the same biological function or activity

as the mature polypeptide of the present invention encoded by a coding sequence which includes the DNA of Figures 1-13 or the deposited cDNA(s).

Alternatively, the polynucleotide may have at least 10 or 20 bases, preferably at least 30 bases, and more preferably at least 50 bases which hybridize to a polynucleotide of the present invention and which has an identity thereto, as hereinabove described, and which may or may not retain activity. For example, such polynucleotides may be employed as probes for polynucleotides, for example, for recovery of the polynucleotide or as a diagnostic probe or as a PCR primer.

Thus, the present invention is directed to polynucleotides having at least a 70% identity, preferably at least 90% and more preferably at least 95% identity to a polynucleotide which encodes the mature polypeptide encoded by a human gene which includes the DNA of one of Figures 1-13 as well as fragments thereof, which fragments have at least 30 bases and preferably at least 50 bases and to polypeptides encoded by such polynucleotides.

The partial sequences are specific tags for messenger RNA molecules. The complete sequence of that messenger RNA, in the form of cDNA, can be determined using the partial sequence as a probe to identify a cDNA clone corresponding to a full-length transcript, followed by sequencing of that clone. The partial cDNA clone can also be used as a probe to identify a genomic clone or clones that contain the complete gene including regulatory and promoter regions, exons, and introns.

The partial sequences of Figures 1-7 and 10-13 may be used to identify the corresponding full length gene from which they were derived. The partial sequences can be nick-translated or end-labelled with ^{32}P using polynucleotide kinase using labelling methods known to those with skill in the art (Basic Methods in Molecular Biology, L.G. Davis, M.D.

Dibner, and J.F. Battey, ed., Elsevier Press, NY, 1986). A lambda library prepared from human colon tissue can be directly screened with the labelled sequences of interest or the library can be converted en masse to pBluescript (Stratagene Cloning Systems, La Jolla, CA 92037) to facilitate bacterial colony screening. Regarding pBluescript, see Sambrook et al., Molecular Cloning-A Laboratory Manual, Cold Spring Harbor Laboratory Press (1989), pg. 1.20. Both methods are well known in the art. Briefly, filters with bacterial colonies containing the library in pBluescript or bacterial lawns containing lambda plaques are denatured and the DNA is fixed to the filters. The filters are hybridized with the labelled probe using hybridization conditions described by Davis et al., supra. The partial sequences, cloned into lambda or pBluescript, can be used as positive controls to assess background binding and to adjust the hybridization and washing stringencies necessary for accurate clone identification. The resulting autoradiograms are compared to duplicate plates of colonies or plaques; each exposed spot corresponds to a positive colony or plaque. The colonies or plaques are selected, expanded and the DNA is isolated from the colonies for further analysis and sequencing.

Positive cDNA clones are analyzed to determine the amount of additional sequence they contain using PCR with one primer from the partial sequence and the other primer from the vector. Clones with a larger vector-insert PCR product than the original partial sequence are analyzed by restriction digestion and DNA sequencing to determine whether they contain an insert of the same size or similar as the mRNA size determined from Northern blot Analysis.

Once one or more overlapping cDNA clones are identified, the complete sequence of the clones can be determined. The preferred method is to use exonuclease III digestion (McCombie, W.R, Kirkness, E., Fleming, J.T., Kerlavage, A.R.,

Iovannisci, D.M., and Martin-Gallardo, R., Methods, 3:33-40, 1991). A series of deletion clones are generated, each of which is sequenced. The resulting overlapping sequences are assembled into a single contiguous sequence of high redundancy (usually three to five overlapping sequences at each nucleotide position), resulting in a highly accurate final sequence.

The DNA sequences (as well as the corresponding RNA sequences) also include sequences which are or contain a DNA sequence identical to one contained in and isolatable from ATCC Deposit No. 97102, deposited March 20, 1995, and fragments or portions of the isolated DNA sequences (and corresponding RNA sequences), as well as DNA (RNA) sequences encoding the same polypeptide.

The deposit(s) referred to herein will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Micro-organisms for purposes of Patent Procedure. These deposits are provided merely as convenience to those of skill in the art and are not an admission that a deposit is required under 35 U.S.C. §112. The sequence of the polynucleotides contained in the deposited materials, as well as the amino acid sequence of the polypeptides encoded thereby, are incorporated herein by reference and are controlling in the event of any conflict with any description of sequences herein. A license may be required to make, use or sell the deposited materials, and no such license is hereby granted.

The present invention further relates to polynucleotides which have at least 10 bases, preferably at least 20 bases, and may have 30 or more bases, which polynucleotides are hybridizable to and have at least a 70% identity to RNA (and DNA which corresponds to such RNA) transcribed from a human gene whose coding portion includes DNA as hereinabove described.

Thus, the polynucleotide sequences which hybridize as described above may be used to hybridize to and detect the expression of the human genes to which they correspond for use in diagnostic assays as hereinafter described.

In accordance with still another aspect of the present invention there are provided diagnostic assays for detecting micrometastases of colon cancer in a host. While applicant does not wish to limit the reasoning of the present invention to any specific scientific theory, it is believed that the presence of active transcription of a colon specific gene of the present invention in cells of the host, other than those derived from the colon, is indicative of colon cancer metastases. This is true because, while the colon specific genes are found in all cells of the body, their transcription to mRNA, cDNA and expression products is primarily limited to the colon in non-diseased individuals. However, if colon cancer is present, colon cancer cells migrate from the cancer to other cells, such that these other cells are now actively transcribing and expressing a colon specific gene at a greater level than is normally found in non-diseased individuals, i.e., transcription is higher than found in non-colon tissues in healthy individuals. It is the detection of this enhanced transcription or enhanced protein expression in cells, other than those derived from the colon, which is indicative of metastases of colon cancer.

In one example of such a diagnostic assay, an RNA sequence in a sample derived from a tissue other than the colon is detected by hybridization to a probe. The sample contains a nucleic acid or a mixture of nucleic acids, at least one of which is suspected of containing a human colon specific gene or fragment thereof of the present invention which is transcribed and expressed in such tissue. Thus, for example, in a form of an assay for determining the presence of a specific RNA in cells, initially RNA is isolated from the cells.

A sample may be obtained from cells derived from tissue other than from the colon including but not limited to blood, urine, saliva, tissue biopsy and autopsy material. The use of such methods for detecting enhanced transcription to mRNA from a human colon specific gene of the present invention or fragment thereof in a sample obtained from cells derived from other than the colon is well within the scope of those skilled in the art from the teachings herein.

The isolation of mRNA comprises isolating total cellular RNA by disrupting a cell and performing differential centrifugation. Once the total RNA is isolated, mRNA is isolated by making use of the adenine nucleotide residues known to those skilled in the art as a poly(A) tail found on virtually every eukaryotic mRNA molecule at the 3' end thereof. Oligonucleotides composed of only deoxythymidine [oligo(dT)] are linked to cellulose and the oligo(dT)-cellulose packed into small columns. When a preparation of total cellular RNA is passed through such a column, the mRNA molecules bind to the oligo(dT) by the poly(A)tails while the rest of the RNA flows through the column. The bound mRNAs are then eluted from the column and collected.

One example of detecting isolated mRNA transcribed from a colon specific gene of the present invention comprises screening the collected mRNAs with the gene specific oligonucleotide probes, as hereinabove described.

It is also appreciated that such probes can be and are preferably labeled with an analytically detectable reagent to facilitate identification of the probe. Useful reagents include but are not limited to radioactivity, fluorescent dyes or enzymes capable of catalyzing the formation of a detectable product.

An example of detecting a polynucleotide complementary to the mRNA sequence (cDNA) utilizes the polymerase chain reaction (PCR) in conjunction with reverse transcriptase. PCR is a very powerful method for the specific amplification

of DNA or RNA stretches (Saiki *et al.*, Nature, 234:163-166 (1986)). One application of this technology is in nucleic acid probe technology to bring up nucleic acid sequences present in low copy numbers to a detectable level. Numerous diagnostic and scientific applications of this method have been described by H.A. Erlich (ed.) in PCR Technology-Principles and Applications for DNA Amplification, Stockton Press, USA, 1989, and by M.A. Inis (ed.) in PCR Protocols, Academic Press, San Diego, USA, 1990.

RT-PCR is a combination of PCR with the reverse transcriptase enzyme. Reverse transcriptase is an enzyme which produces cDNA molecules from corresponding mRNA molecules. This is important since PCR amplifies nucleic acid molecules, particularly DNA, and this DNA may be produced from the mRNA isolated from a sample derived from the host.

A specific example of an RT-PCR diagnostic assay involves removing a sample from a tissue of a host. Such a sample will be from a tissue, other than the colon, for example, blood. Therefore, an example of such a diagnostic assay comprises whole blood gradient isolation of nucleated cells, total RNA extraction, RT-PCR of total RNA and agarose gel electrophoresis of PCR products. The PCR products comprise cDNA complementary to RNA transcribed from one or more colon specific genes of the present invention or fragments thereof. More particularly, a blood sample is obtained and the whole blood is combined with an equal volume of phosphate buffered saline, centrifuged and the lymphocyte and granulocyte layer is carefully aspirated and rediluted in phosphate buffered saline and centrifuged again. The supernate is discarded and the pellet containing nucleated cells is used for RNA extraction using the RNazole B method as described by the manufacturer (Tel-Test Inc., Friendswood, TX).

Oligonucleotide primers and probes are prepared with high specificity to the DNA sequences of the present invention. The probes are at least 10 base pairs in length, preferably at least 30 base pairs in length and most preferably at least 50 base pairs in length or more. The reverse transcriptase reaction and PCR amplification are performed sequentially without interruption. Taq polymerase is used during PCR and the PCR products are concentrated and the entire sample is run on a Tris-borate-EDTA agarose gel containing ethidium bromide.

Another aspect of the present invention relates to assays which detect the presence of an altered level of the expression products of the colon specific genes of the present invention. Thus, for example, such an assay involves detection of the polypeptides of the present invention or fragments thereof.

In accordance with another aspect of the present invention, there is provided a method of diagnosing a disorder of the colon, for example colon cancer, by determining altered levels of the colon specific polypeptides of the present invention in a biological sample, derived from tissue other than from the colon. Elevated levels of the colon specific polypeptides of the present invention, excluding CSG7 and CSG10, indicates active transcription and expression of the corresponding colon specific gene product. Assays used to detect levels of a colon specific gene polypeptide in a sample derived from a host are well-known to those skilled in the art and include radioimmunoassays, competitive-binding assays, Western blot analysis, ELISA assays and "sandwich" assays. A biological sample may include, but is not limited to, tissue extracts, cell samples or biological fluids, however, in accordance with the present invention, a biological sample specifically does not include tissue or cells of the colon.

An ELISA assay (Coligan, et al., Current Protocols in Immunology, 1(2), Chapter 6, 1991) initially comprises preparing an antibody specific to a colon specific polypeptide of the present invention, preferably a monoclonal antibody. In addition, a reporter antibody is prepared against the monoclonal antibody. To the reporter antibody is attached a detectable reagent such as radioactivity, fluorescence or, in this example, a horseradish peroxidase enzyme. A sample is removed from a host and incubated on a solid support, e.g., a polystyrene dish, that binds the proteins in the sample. Any free protein binding sites on the dish are then covered by incubating with a non-specific protein, such as BSA. Next, the monoclonal antibody is incubated in the dish during which time the monoclonal antibodies attach to the colon specific polypeptide attached to the polystyrene dish. All unbound monoclonal antibody is washed out with buffer. The reporter antibody linked to horseradish peroxidase is now placed in the dish resulting in binding of the reporter antibody to any monoclonal antibody bound to the colon specific gene polypeptide. Unattached reporter antibody is then washed out. Peroxidase substrates are then added to the dish and the amount of color developed in a given time period is a measurement of the amount of the colon specific polypeptide present in a given volume of patient sample when compared against a standard curve.

A competition assay may be employed where antibodies specific to a colon specific polypeptide are attached to a solid support. The colon specific polypeptide is then labeled and the labeled polypeptide a sample derived from the host are passed over the solid support and the amount of label detected, for example, by liquid scintillation chromatography, can be correlated to a quantity of the colon specific polypeptide in the sample.

A "sandwich" assay is similar to an ELISA assay. In a "sandwich" assay, colon specific polypeptides are passed over

a solid support and bind to antibody attached to the solid support. A second antibody is then bound to the colon specific polypeptide. A third antibody which is labeled and is specific to the second antibody, is then passed over the solid support and binds to the second antibody and an amount can then be quantified.

In alternative methods, labeled antibodies to a colon specific polypeptide are used. In a one-step assay, the target molecule, if it is present, is immobilized and incubated with a labeled antibody. The labeled antibody binds to the immobilized target molecule. After washing to remove the unbound molecules, the sample is assayed for the presence of the label. In a two-step assay, immobilized target molecule is incubated with an unlabeled antibody. The target molecule-labeled antibody complex, if present, is then bound to a second, labeled antibody that is specific for the unlabeled antibody. The sample is washed and assayed for the presence of the label.

The choice of marker used to label the antibodies will vary depending upon the application. However, the choice of marker is readily determinable to one skilled in the art. These labeled antibodies may be used in immunoassays as well as in histological applications to detect the presence of the proteins. The labeled antibodies may be polyclonal or monoclonal.

The presence of active transcription, which is greater than that normally found, of the colon specific genes in cells other than from the colon, by the presence of an altered level of mRNA, cDNA or expression products is an important indication of the presence of a colon cancer which has metastasized, since colon cancer cells are migrating from the colon into the general circulation. Accordingly, this phenomenon may have important clinical implications since the method of treating a localized, as opposed to a metastasized, tumor is entirely different.

The assays described above may also be used to test whether bone marrow preserved before chemotherapy is contaminated with micrometastases of a colon cancer cell. In the assay, blood cells from the bone marrow are isolated and treated as described above, this method allows one to determine whether preserved bone marrow is still suitable for transplantation after chemotherapy.

The present invention further relates to mature polypeptides as well as fragments, analogs and derivatives of such polypeptide.

The terms "fragment," "derivative" and "analog" when referring to the polypeptides encoded by the genes of the invention means a polypeptide which retains essentially the same biological function or activity as such polypeptide. Thus, an analog includes a proprotein which can be activated by cleavage of the proprotein portion to produce an active mature polypeptide.

The polypeptides of the present invention may be recombinant polypeptides, natural polypeptides or synthetic polypeptides, preferably recombinant polypeptides.

The fragment, derivative or analog of the polypeptides encoded by the genes of the invention may be (i) one in which one or more of the amino acid residues are substituted with a conserved or non-conserved amino acid residue (preferably a conserved amino acid residue) and such substituted amino acid residue may or may not be one encoded by the genetic code, or (ii) one in which one or more of the amino acid residues includes a substituent group, or (iii) one in which the polypeptide is fused with another compound, such as a compound to increase the half-life of the polypeptide (for example, polyethylene glycol), or (iv) one in which the additional amino acids are fused to the polypeptide, such as a leader or secretory sequence or a sequence which is

employed for purification of the mature polypeptide or a proprotein sequence. Such fragments, derivatives and analogs are deemed to be within the scope of those skilled in the art from the teachings herein.

The polypeptides and polynucleotides of the present invention are preferably provided in an isolated form, and preferably are purified to homogeneity.

The term "isolated" means that the material is removed from its original environment (e.g., the natural environment if it is naturally occurring). For example, a naturally-occurring polynucleotide or polypeptide present in a living animal is not isolated, but the same polynucleotide or polypeptide, separated from some or all of the coexisting materials in the natural system, is isolated. Such polynucleotides could be part of a vector and/or such polynucleotides or polypeptides could be part of a composition, and still be isolated in that such vector or composition is not part of its natural environment.

The polypeptides of the present invention include the polypeptides of Figures 8 and 9 (in particular the mature polypeptides) as well as polypeptides which have at least 70% similarity (preferably at least a 70% identity) to the polypeptides of Figures 8 and 9 and more preferably at least a 90% similarity (more preferably at least a 90% identity) to the polypeptides of Figures 8 and 9 and still more preferably at least a 95% similarity (still more preferably at least 95% identity) to the polypeptides of Figures 8 and 9 and also include portions of such polypeptides with such portion of the polypeptide generally containing at least 30 amino acids and more preferably at least 50 amino acids.

As known in the art "similarity" between two polypeptides is determined by comparing the amino acid sequence and its conserved amino acid substitutes of one polypeptide to the sequence of a second polypeptide.

Fragments or portions of the polypeptides of the present invention may be employed for producing the corresponding full-length polypeptide by peptide synthesis; therefore, the fragments may be employed as intermediates for producing the full-length polypeptides. Fragments or portions of the polynucleotides of the present invention may be used to synthesize full-length polynucleotides of the present invention.

The present invention also relates to vectors which include polynucleotides of the present invention, host cells which are genetically engineered with vectors of the invention and the production of polypeptides of the invention by recombinant techniques.

Host cells are genetically engineered (transduced or transformed or transfected) with the vectors of this invention which may be, for example, a cloning vector or an expression vector. The vector may be, for example, in the form of a plasmid, a viral particle, a phage, etc. The engineered host cells can be cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants or amplifying the colon specific genes. The culture conditions, such as temperature, pH and the like, are those previously used with the host cell selected for expression, and will be apparent to those of ordinarily skill in the art.

The polynucleotides of the present invention may be employed for producing polypeptides by recombinant techniques. Thus, for example, the polynucleotide may be included in any one of a variety of expression vectors for expressing a polypeptide. Such vectors include chromosomal, nonchromosomal and synthetic DNA sequences, e.g., derivatives of SV40; bacterial plasmids; phage DNA; baculovirus; yeast plasmids; vectors derived from combinations of plasmids and phage DNA, viral DNA such as vaccinia, adenovirus, fowl pox virus, and pseudorabies.

However, any other vector may be used as long as it is replicable and viable in the host.

The appropriate DNA sequence may be inserted into the vector by a variety of procedures. In general, the DNA sequence is inserted into an appropriate restriction endonuclease site(s) by procedures known in the art. Such procedures and others are deemed to be within the scope of those skilled in the art.

The DNA sequence in the expression vector is operatively linked to an appropriate expression control sequence(s) (promoter) to direct mRNA synthesis. As representative examples of such promoters, there may be mentioned: LTR or SV40 promoter, the E. coli. lac or trp, the phage lambda P_L promoter and other promoters known to control expression of genes in prokaryotic or eukaryotic cells or their viruses. The expression vector also contains a ribosome binding site for translation initiation and a transcription terminator. The vector may also include appropriate sequences for amplifying expression.

In addition, the expression vectors preferably contain one or more selectable marker genes to provide a phenotypic trait for selection of transformed host cells such as dihydrofolate reductase or neomycin resistance for eukaryotic cell culture, or such as tetracycline or ampicillin resistance in E. coli.

The vector containing the appropriate DNA sequence as hereinabove described, as well as an appropriate promoter or control sequence, may be employed to transform an appropriate host to permit the host to express the protein.

As representative examples of appropriate hosts, there may be mentioned: bacterial cells, such as E. coli, Streptomyces, Salmonella typhimurium; fungal cells, such as yeast; insect cells such as Drosophila S2 and Spodoptera Sf9; animal cells such as CHO, COS or Bowes melanoma; adenoviruses; plant cells, etc. The selection of an

appropriate host is deemed to be within the scope of those skilled in the art from the teachings herein.

More particularly, the present invention also includes recombinant constructs comprising one or more of the sequences as broadly described above. The constructs comprise a vector, such as a plasmid or viral vector, into which a sequence of the invention has been inserted, in a forward or reverse orientation. In a preferred aspect of this embodiment, the construct further comprises regulatory sequences, including, for example, a promoter, operably linked to the sequence. Large numbers of suitable vectors and promoters are known to those of skill in the art, and are commercially available. The following vectors are provided by way of example. Bacterial: pQE70, pQE60, pQE-9 (Qiagen), pBS, pD10, phagescript, psiX174, pbluescript SK, pBSKS, pNH8A, pNH16a, pNH18A, pNH46A (Stratagene); ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia). Eukaryotic: pWLNEO, pSV2CAT, pOG44, pXT1, pSG (Stratagene) pSVK3, pBPV, pMSG, pSVL (Pharmacia). However, any other plasmid or vector may be used as long as they are replicable and viable in the host.

Promoter regions can be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers. Two appropriate vectors are pKK232-8 and pCM7. Particular named bacterial promoters include lacI, lacZ, T3, T7, gpt, lambda P_R, P_L and trp. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art.

In a further embodiment, the present invention relates to host cells containing the above-described constructs. The host cell can be a higher eukaryotic cell, such as a mammalian cell, or a lower eukaryotic cell, such as a yeast

cell, or the host cell can be a prokaryotic cell, such as a bacterial cell. Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE-Dextran mediated transfection, or electroporation (Davis, L., Dibner, M., Battey, I., Basic Methods in Molecular Biology, (1986)).

The constructs in host cells can be used in a conventional manner to produce the gene product encoded by the recombinant sequence. Alternatively, the polypeptides of the invention can be synthetically produced by conventional peptide synthesizers.

Proteins can be expressed in mammalian cells, yeast, bacteria, or other cells under the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described by Sambrook, et al., Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor, N.Y., (1989), the disclosure of which is hereby incorporated by reference.

Transcription of the DNA encoding the polypeptides of the present invention by higher eukaryotes is increased by inserting an enhancer sequence into the vector. Enhancers are cis-acting elements of DNA, usually about from 10 to 300 bp that act on a promoter to increase its transcription. Examples including the SV40 enhancer on the late side of the replication origin bp 100 to 270, a cytomegalovirus early promoter enhancer, the polyoma enhancer on the late side of the replication origin, and adenovirus enhancers.

Generally, recombinant expression vectors will include origins of replication and selectable markers permitting transformation of the host cell, e.g., the ampicillin resistance gene of E. coli and S. cerevisiae TRP1 gene, and a promoter derived from a highly-expressed gene to direct

transcription of a downstream structural sequence. Such promoters can be derived from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK), α-factor, acid phosphatase, or heat shock proteins, among others. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequences. Optionally, the heterologous sequence can encode a fusion protein including an N-terminal identification peptide imparting desired characteristics, e.g., stabilization or simplified purification of expressed recombinant product.

Useful expression vectors for bacterial use are constructed by inserting a structural DNA sequence encoding a desired protein together with suitable translation initiation and termination signals in operable reading frame with a functional promoter. The vector will comprise one or more phenotypic selectable markers and an origin of replication to ensure maintenance of the vector and to, if desirable, provide amplification within the host. Suitable prokaryotic hosts for transformation include E. coli, Bacillus subtilis, Salmonella typhimurium and various species within the genera Pseudomonas, Streptomyces, and Staphylococcus, although others may also be employed as a matter of choice.

As a representative but nonlimiting example, useful expression vectors for bacterial use can comprise a selectable marker and bacterial origin of replication derived from commercially available plasmids comprising genetic elements of the well known cloning vector pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (Pharmacia Fine Chemicals, Uppsala, Sweden) and GEM1 (Promega Biotec, Madison, WI, USA). These pBR322 "backbone" sections are combined with an appropriate promoter and the structural sequence to be expressed.

Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the

selected promoter is induced by appropriate means (e.g., temperature shift or chemical induction) and cells are cultured for an additional period.

Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification.

Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents, such methods are well know to those skilled in the art.

Various mammalian cell culture systems can also be employed to express recombinant protein. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts, described by Gluzman, Cell, 23:175 (1981), and other cell lines capable of expressing a compatible vector, for example, the C127, 3T3, CHO, HeLa and BHK cell lines. Mammalian expression vectors will comprise an origin of replication, a suitable promoter and enhancer, and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking nontranscribed sequences. DNA sequences derived from the SV40 splice, and polyadenylation sites may be used to provide the required nontranscribed genetic elements.

The colon specific gene polypeptides can be recovered and purified from recombinant cell cultures by methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Protein refolding steps can be used, as necessary, in completing configuration of the mature protein. Finally, high performance liquid

chromatography (HPLC) can be employed for final purification steps.

The polynucleotides of the present invention may have the coding sequence fused in frame to a marker sequence which allows for purification of the polypeptide of the present invention. An example of a marker sequence is a hexahistidine tag which may be supplied by a vector, preferably a pQE-9 vector, which provides for purification of the polypeptide fused to the marker in the case of a bacterial host, or, for example, the marker sequence may be a hemagglutinin (HA) tag when a mammalian host, e.g. COS-7 cells, is used. The HA tag corresponds to an epitope derived from the influenza hemagglutinin protein (Wilson, I., et al., Cell, 37:767 (1984)).

The polypeptides of the present invention may be a naturally purified product, or a product of chemical synthetic procedures, or produced by recombinant techniques from a prokaryotic or eukaryotic host (for example, by bacterial, yeast, higher plant, insect and mammalian cells in culture). Depending upon the host employed in a recombinant production procedure, the polypeptides of the present invention may be glycosylated or may be non-glycosylated. Polypeptides of the invention may also include an initial methionine amino acid residue.

In accordance with another aspect of the present invention there are provided assays which may be used to screen for therapeutics to inhibit the action of the colon specific genes or colon specific proteins of the present invention, excluding CSG7 and CSG10. One assay takes advantage of the reductase function of these proteins. The present invention discloses methods for selecting a therapeutic which forms a complex with colon specific gene proteins with sufficient affinity to prevent their biological action. The methods include various assays, including competitive assays where the proteins are immobilized to a

support, and are contacted with a natural substrate and a labeled therapeutic either simultaneously or in either consecutive order, and determining whether the therapeutic effectively competes with the natural substrate in a manner sufficient to prevent binding of the protein to its substrate.

In another embodiment, the substrate is immobilized to a support, and is contacted with both a labeled colon specific polypeptide and a therapeutic (or unlabeled proteins and a labeled therapeutic), and it is determined whether the amount of the colon specific polypeptide bound to the substrate is reduced in comparison to the assay without the therapeutic added. The colon specific polypeptide may be labeled with antibodies.

In another example of such a screening assay, there is provided a mammalian cell or membrane preparation expressing a colon specific polypeptide of the present invention incubated with elements which undergo simultaneous oxidation and reduction, for example hydrogen and oxygen which together form water, wherein the hydrogen could be labeled by radioactivity, e.g., tritium, in the presence of the compound to be screened under conditions favoring the oxidation reduction reaction where hydrogen and oxygen form water. The ability of the compound to enhance or block this interaction could then be measured.

Potential therapeutic compounds include antibodies and anti-idiotypic antibodies as described above, or in some cases, an oligonucleotide, which binds to the polypeptide.

Another example is an antisense construct prepared using antisense technology, which is directed to a colon specific polynucleotide to prevent transcription. Antisense technology can be used to control gene expression through triple-helix formation or antisense DNA or RNA, both of which methods are based on binding of a polynucleotide to DNA or RNA. For example, the 5′ coding portion of the

polynucleotide sequence, which encodes for the mature polypeptides of the present invention, is used to design an antisense RNA oligonucleotide of from about 10 to 40 base pairs in length. A DNA oligonucleotide is designed to be complementary to a region of the gene involved in transcription (triple helix -see Lee et al., Nucl. Acids Res., 6:3073 (1979); Cooney et al, Science, 241:456 (1988); and Dervan et al., Science, 251: 1360 (1991)), thereby preventing transcription and the production of a colon specific polynucleotide. The antisense RNA oligonucleotide hybridizes to the mRNA *in vivo* and blocks translation of the mRNA molecule into the colon specific genes polypeptide (antisense - Okano, J. Neurochem., 56:560 (1991); Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988)). The oligonucleotides described above can also be delivered to cells such that the antisense RNA or DNA may be expressed *in vivo* to inhibit production of the colon specific polypeptides.

Another example is a small molecule which binds to and occupies the active site of the colon specific polypeptide thereby making the active site inaccessible to substrate such that normal biological activity is prevented. Examples of small molecules include but are not limited to small peptides or peptide-like molecules.

These compounds may be employed to treat colon cancer, since they interact with the function of colon specific polypeptides in a manner sufficient to inhibit natural function which is necessary for the viability of colon cancer cells. The compounds may be employed in a composition with a pharmaceutically acceptable carrier, e.g., as hereinafter described.

The compounds of the present invention may be employed in combination with a suitable pharmaceutical carrier. Such compositions comprise a therapeutically effective amount of

the polypeptide, and a pharmaceutically acceptable carrier or excipient. Such a carrier includes but is not limited to saline, buffered saline, dextrose, water, glycerol, ethanol, and combinations thereof. The formulation should suit the mode of administration.

The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration. In addition, the pharmaceutical compositions may be employed in conjunction with other therapeutic compounds.

The pharmaceutical compositions may be administered in a convenient manner such as by the oral, topical, intravenous, intraperitoneal, intramuscular, subcutaneous, intranasal, intra-anal or intradermal routes. The pharmaceutical compositions are administered in an amount which is effective for treating and/or prophylaxis of the specific indication. In general, they are administered in an amount of at least about 10 μg/kg body weight and in most cases they will be administered in an amount not in excess of about 8 mg/Kg body weight per day. In most cases, the dosage is from about 10 μg/kg to about 1 mg/kg body weight daily, taking into account the routes of administration, symptoms, etc.

The colon specific genes and compounds which are polypeptides may also be employed in accordance with the present invention by expression of such polypeptides *in vivo*, which is often referred to as "gene therapy."

Thus, for example, cells from a patient may be engineered with a polynucleotide (DNA or RNA) encoding a polypeptide *ex vivo*, with the engineered cells then being

provided to a patient to be treated with the polypeptide. Such methods are well-known in the art. For example, cells may be engineered by procedures known in the art by use of a retroviral particle containing RNA encoding a polypeptide of the present invention.

Similarly, cells may be engineered *in vivo* for expression of a polypeptide *in vivo* by, for example, procedures known in the art. As known in the art, a producer cell for producing a retroviral particle containing RNA encoding a polypeptide of the present invention may be administered to a patient for engineering cells *in vivo* and expression of the polypeptide *in vivo*. These and other methods for administering a polypeptide of the present invention by such method should be apparent to those skilled in the art from the teachings of the present invention. For example, the expression vehicle for engineering cells may be other than a retrovirus, for example, an adenovirus which may be used to engineer cells in vivo after combination with a suitable delivery vehicle.

Retroviruses from which the retroviral plasmid vectors hereinabove mentioned may be derived include, but are not limited to, Moloney Murine Leukemia Virus, spleen necrosis virus, retroviruses such as Rous Sarcoma Virus, Harvey Sarcoma Virus, avian leukosis virus, gibbon ape leukemia virus, human immunodeficiency virus, adenovirus, Myeloproliferative Sarcoma Virus, and mammary tumor virus. In one embodiment, the retroviral plasmid vector is derived from Moloney Murine Leukemia Virus.

The vector includes one or more promoters. Suitable promoters which may be employed include, but are not limited to, the retroviral LTR; the SV40 promoter; and the human cytomegalovirus (CMV) promoter described in Miller, et al., Biotechniques, Vol. 7, No. 9, 980-990 (1989), or any other promoter (e.g., cellular promoters such as eukaryotic cellular promoters including, but not limited to, the

histone, pol III, and β-actin promoters). Other viral promoters which may be employed include, but are not limited to, adenovirus promoters, thymidine kinase (TK) promoters, and B19 parvovirus promoters. The selection of a suitable promoter will be apparent to those skilled in the art from the teachings contained herein.

The nucleic acid sequence encoding the polypeptide of the present invention is under the control of a suitable promoter. Suitable promoters which may be employed include, but are not limited to, adenoviral promoters, such as the adenoviral major late promoter; or heterologous promoters, such as the cytomegalovirus (CMV) promoter; the respiratory syncytial virus (RSV) promoter; inducible promoters, such as the MMT promoter, the metallothionein promoter; heat shock promoters; the albumin promoter; the ApoAI promoter; human globin promoters; viral thymidine kinase promoters, such as the Herpes Simplex thymidine kinase promoter; retroviral LTRs (including the modified retroviral LTRs hereinabove described); the β-actin promoter; and human growth hormone promoters. The promoter also may be the native promoter which controls the genes encoding the polypeptides.

The retroviral plasmid vector is employed to transduce packaging cell lines to form producer cell lines. Examples of packaging cells which may be transfected include, but are not limited to, the PE501, PA317, ψ-2, ψ-AM, PA12, T19-14X, VT-19-17-H2, ψCRE, ψCRIP, GP+E-86, GP+envAm12, and DAN cell lines as described in Miller, Human Gene Therapy, Vol. 1, pgs. 5-14 (1990), which is incorporated herein by reference in its entirety. The vector may transduce the packaging cells through any means known in the art. Such means include, but are not limited to, electroporation, the use of liposomes, and $CaPO_4$ precipitation. In one alternative, the retroviral plasmid vector may be encapsulated into a liposome, or coupled to a lipid, and then administered to a host.

The producer cell line generates infectious retroviral vector particles which include the nucleic acid sequence(s) encoding the polypeptides. Such retroviral vector particles then may be employed, to transduce eukaryotic cells, either *in vitro* or *in vivo*. The transduced eukaryotic cells will express the nucleic acid sequence(s) encoding the polypeptide. Eukaryotic cells which may be transduced include, but are not limited to, embryonic stem cells, embryonic carcinoma cells, as well as hematopoietic stem cells, hepatocytes, fibroblasts, myoblasts, keratinocytes, endothelial cells, and bronchial epithelial cells.

This invention is also related to the use of a colon specific genes of the present invention as a diagnostic. For example, some diseases result from inherited defective genes. The colon specific genes, CSG7 and CSG10, for example, have been found to have a reduced expression in colon cancer cells as compared to that in normal cells. Further, the remaining colon specific genes of the present invention are overexpressed in colon cancer. Accordingly, a mutation in these genes allows a detection of colon disorders, for example, colon cancer. A mutation in a colon specific gene of the present invention at the DNA level may be detected by a variety of techniques. Nucleic acids used for diagnosis (genomic DNA, mRNA, etc.) may be obtained from a patient's cells, other than from the colon, such as from blood, urine, saliva, tissue biopsy and autopsy material. The genomic DNA may be used directly for detection or may be amplified enzymatically by using PCR (Saiki, et al., Nature, 324:163-166 (1986)) prior to analysis. RNA or cDNA may also be used for the same purpose. As an example, PCR primers complementary to the nucleic acid of the instant invention can be used to identify and analyze mutations in a colon specific polynucleotide of the present invention. For example, deletions and insertions can be detected by a change in size of the amplified product in comparison to the normal

genotype. Point mutations can be identified by hybridizing amplified DNA to radiolabelled colon specific RNA or, alternatively, radiolabelled antisense DNA sequences.

Another well-established method for screening for mutations in particular segments of DNA after PCR amplification is single-strand conformation polymorphism (SSCP) analysis. PCR products are prepared for SSCP by ten cycles of reamplification to incorporate ^{32}P-dCTP, digested with an appropriate restriction enzyme to generate 200-300 bp fragments, and denatured by heating to 85°C for 5 min. and then plunged into ice. Electrophoresis is then carried out in a nondenaturing gel (5% glycerol, 5% acrylamide) (Glavac, D. and Dean, M., Human Mutation, 2:404-414 (1993)).

Sequence differences between the reference gene and "mutants" may be revealed by the direct DNA sequencing method. In addition, cloned DNA segments may be used as probes to detect specific DNA segments. The sensitivity of this method is greatly enhanced when combined with PCR. For example, a sequencing primer is used with double-stranded PCR product or a single-stranded template molecule generated by a modified PCR. The sequence determination is performed by conventional procedures with radiolabeled nucleotides or by automatic sequencing procedures with fluorescent-tags.

Genetic testing based on DNA sequence differences may be achieved by detection of alteration in electrophoretic mobility of DNA fragments and gels with or without denaturing agents. Small sequence deletions and insertions can be visualized by high-resolution gel electrophoresis. DNA fragments of different sequences may be distinguished on denaturing formamide gradient gels in which the mobilities of different DNA fragments are retarded in the gel at different positions according to their specific melting or partial melting temperatures (see, e.g., Myers, et al., Science, 230:1242 (1985)). In addition, sequence alterations, in

particular small deletions, may be detected as changes in the migration pattern of DNA.

Sequence changes at specific locations may also be revealed by nuclease protection assays, such as Rnase and S1 protection or the chemical cleavage method (e.g., Cotton, et al., PNAS, USA, 85:4397-4401 (1985)).

Thus, the detection of the specific DNA sequence may be achieved by methods such as hybridization, RNase protection, chemical cleavage, direct DNA sequencing, or the use of restriction enzymes (e.g., Restriction Fragment Length Polymorphisms (RFLP)) and Southern blotting.

The sequences of the present invention are also valuable for chromosome identification. The sequence is specifically targeted to and can hybridize with a particular location on an individual human chromosome. Moreover, there is a current need for identifying particular sites on the chromosome. Few chromosome marking reagents based on actual sequence data (repeat polymorphisms) are presently available for marking chromosomal location. The mapping of DNAs to chromosomes according to the present invention is an important first step in correlating those sequences with genes associated with disease.

Briefly, sequences can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp) from the cDNA. Computer analysis of the 3' untranslated region is used to rapidly select primers that do not span more than one exon in the genomic DNA, thus complicating the amplification process. These primers are then used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the primer will yield an amplified fragment.

PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular DNA to a particular chromosome. Using the present invention with the same oligonucleotide primers, sublocalization can be achieved with panels of

fragments from specific chromosomes or pools of large genomic clones in an analogous manner. Other mapping strategies that can similarly be used to map to its chromosome include *in situ* hybridization, prescreening with labeled flow-sorted chromosomes and preselection by hybridization to construct chromosome specific-cDNA libraries.

Fluorescence *in situ* hybridization (FISH) of a cDNA clone to a metaphase chromosomal spread can be used to provide a precise chromosomal location in one step. This technique can be used with cDNA as short as 50 or 60 bases. For a review of this technique, see Verma et al., Human Chromosomes: a Manual of Basic Techniques, Pergamon Press, New York (1988).

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such data are found, for example, in V. McKusick, Mendelian Inheritance in Man (available on line through Johns Hopkins University Welch Medical Library). The relationship between genes and diseases that have been mapped to the same chromosomal region are then identified through linkage analysis (coinheritance of physically adjacent genes).

Next, it is necessary to determine the differences in the cDNA or genomic sequence between affected and unaffected individuals. If a mutation is observed in some or all of the affected individuals but not in any normal individuals, then the mutation is likely to be the causative agent of the disease.

With current resolution of physical mapping and genetic mapping techniques, a cDNA precisely localized to a chromosomal region associated with the disease could be one of between 50 and 500 potential causative genes. (This assumes 1 megabase mapping resolution and one gene per 20 kb).

The polypeptides, their fragments or other derivatives, or analogs thereof, or cells expressing them can be used as an immunogen to produce antibodies thereto. These antibodies can be, for example, polyclonal or monoclonal antibodies. The present invention also includes chimeric, single chain, and humanized antibodies, as well as Fab fragments, or the product of an Fab expression library. Various procedures known in the art may be used for the production of such antibodies and fragments.

Antibodies generated against the polypeptides corresponding to a sequence of the present invention can be obtained by direct injection of the polypeptides into an animal or by administering the polypeptides to an animal, preferably a nonhuman. The antibody so obtained will then bind the polypeptides itself. In this manner, even a sequence encoding only a fragment of the polypeptides can be used to generate antibodies binding the whole native polypeptides. Such antibodies can then be used to isolate the polypeptide from tissue expressing that polypeptide.

For preparation of monoclonal antibodies, any technique which provides antibodies produced by continuous cell line cultures can be used. Examples include the hybridoma technique (Kohler and Milstein, 1975, Nature, 256:495-497), the trioma technique, the human B-cell hybridoma technique (Kozbor et al., 1983, Immunology Today 4:72), and the EBV-hybridoma technique to produce human monoclonal antibodies (Cole, et al., 1985, in Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc., pp. 77-96).

Techniques described for the production of single chain antibodies (U.S. Patent 4,946,778) can be adapted to produce single chain antibodies to immunogenic polypeptide products of this invention. Transgenic mice may also be used to generate antibodies.

The antibodies may also be employed to target colon cancer cells, for example, in a method of homing interaction

agents which, when contacting colon cancer cells, destroy them. This is true since the antibodies are specific for the colon specific polypeptides of the present invention. A linking of the interaction agent to the antibody would cause the interaction agent to be carried directly to the colon.

Antibodies of this type may also be used to do *in vivo* imaging, for example, by labeling the antibodies to facilitate scanning of the pelvic area and the colon. One method for imaging comprises contacting any cancer cells of the colon to be imaged with an anti-colon specific protein-antibody labeled with a detectable marker. The method is performed under conditions such that the labeled antibody binds to the colon specific polypeptides. In a specific example, the antibodies interact with the colon, for example, colon cancer cells, and fluoresce upon contact such that imaging and visibility of the colon are enhanced to allow a determination of the diseased or non-diseased state of the colon.

The present invention will be further described with reference to the following examples; however, it is to be understood that the present invention is not limited to such examples. All parts or amounts, unless otherwise specified, are by weight.

In order to facilitate understanding of the following examples certain frequently occurring methods and/or terms will be described.

"Plasmids" are designated by a lower case p preceded and/or followed by capital letters and/or numbers. The starting plasmids herein are either commercially available, publicly available on an unrestricted basis, or can be constructed from available plasmids in accord with published procedures. In addition, equivalent plasmids to those described are known in the art and will be apparent to the ordinarily skilled artisan.

"Digestion" of DNA refers to catalytic cleavage of the DNA with a restriction enzyme that acts only at certain sequences in the DNA. The various restriction enzymes used herein are commercially available and their reaction conditions, cofactors and other requirements were used as would be known to the ordinarily skilled artisan. For analytical purposes, typically 1 μg of plasmid or DNA fragment is used with about 2 units of enzyme in about 20 μl of buffer solution. For the purpose of isolating DNA fragments for plasmid construction, typically 5 to 50 μg of DNA are digested with 20 to 250 units of enzyme in a larger volume. Appropriate buffers and substrate amounts for particular restriction enzymes are specified by the manufacturer. Incubation times of about 1 hour at 37°C are ordinarily used, but may vary in accordance with the supplier's instructions. After digestion the reaction is electrophoresed directly on a polyacrylamide gel to isolate the desired fragment.

Size separation of the cleaved fragments is performed using 8 percent polyacrylamide gel described by Goeddel, D. *et al.*, Nucleic Acids Res., 8:4057 (1980).

"Oligonucleotides" refers to either a single stranded polydeoxynucleotide or two complementary polydeoxynucleotide strands which may be chemically synthesized. Such synthetic oligonucleotides have no 5' phosphate and thus will not ligate to another oligonucleotide without adding a phosphate with an ATP in the presence of a kinase. A synthetic oligonucleotide will ligate to a fragment that has not been dephosphorylated.

"Ligation" refers to the process of forming phosphodiester bonds between two double stranded nucleic acid fragments (Maniatis, T., et al., Id., p. 146). Unless otherwise provided, ligation may be accomplished using known buffers and conditions with 10 units of T4 DNA ligase

("ligase") per 0.5 μg of approximately equimolar amounts of the DNA fragments to be ligated.

Unless otherwise stated, transformation was performed as described in the method of Graham, F. and Van der Eb, A., Virology, 52:456-457 (1973).

Example 1

Determination of Transcription of a colon specific gene

To assess the presence or absence of active transcription of a colon specific gene RNA, approximately 6 ml of venous blood is obtained with a standard venipuncture technique using heparinized tubes. Whole blood is mixed with an equal volume of phosphate buffered saline, which is then layered over 8 ml of Ficoll (Pharmacia, Uppsala, Sweden) in a 15-ml polystyrene tube. The gradient is centrifuged at 1800 X g for 20 min at 5°C. The lymphocyte and granulocyte layer (approximately 5 ml) is carefully aspirated and rediluted up to 50 ml with phosphate-buffered saline in a 50-ml tube, which is centrifuged again at 1800 X g for 20 min. at 5°C. The supernatant is discarded and the pellet containing nucleated cells is used for RNA extraction using the RNazole B method as described by the manufacturer (Tel-Test Inc., Friendswood, TX).

To determine the quantity of mRNA from the gene of interest, a probe is designed with an identity to at least portion of the mRNA sequence transcribed from a human gene whose coding portion includes a DNA sequence of one of Figures 1-13. This probe is mixed with the extracted RNA and the mixed DNA and RNA are precipitated with ethanol -70°C for 15 minutes). The pellet is resuspended in hybridization buffer and dissolved. The tubes containing the mixture are incubated in a 72°C water bath for 10-15 mins. to denature the DNA. The tubes are rapidly transferred to a water bath at the desired hybridization temperature. Hybridization temperature depends on the G + C content of the DNA.

Hybridization is done for 3 hrs. 0.3 ml of nuclease-S1 buffer is added and mixed well. 50 μl of 4.0 M ammonium acetate and 0.1 M EDTA is added to stop the reaction. The mixture is extracted with phenol/chloroform and 20 μg of carrier tRNA is added and precipitation is done with an equal volume of isopropanol. The precipitate is dissolved in 40 μl of TE (pH 7.4) and run on an alkaline agarose gel. Following electrophoresis, the RNA is microsequenced to confirm the nucleotide sequence. (See Favaloro, J. et al., Methods Enzymol., 65:718 (1980) for a more detailed review).

Two oligonucleotide primers are employed to amplify the sequence isolated by the above methods. The 5' primer is 20 nucleotides long and the 3' primer is a complimentary sequence for the 3' end of the isolated mRNA. The primers are custom designed according to the isolated mRNA. The reverse transcriptase reaction and PCR amplification are performed sequentially without interruption in a Perkin Elmer 9600 PCR machine (Emeryville, CA). Four hundred ng total RNA in 20 μl diethylpyrocarbonate-treated water are placed in a 65°C water bath for 5 min. and then quickly chilled on ice immediately prior to the addition of PCR reagents. The 50-μl total PCR volume consisted of 2.5 units Taq polymerase (Perkin-Elmer). 2 units avian myeloblastosis virus reverse transcriptase (Boehringer Mannheim, Indianapolis, IN); 200 μM each of dCTP, dATP, dGTP and dTTP (Perkin Elmer); 18 pM each primer, 10 mM Tris-HCl; 50 mM KCl; and 2 mM $MgCl_2$ (Perkin Elmer). PCR conditions are as follows: cycle 1 is 42°C for 15 min then 97°C for 15 s (1 cycle); cycle 2 is 95°C for 1 min. 60°C for 1 min, and 72°C for 30 s (15 cycles); cycle 3 is 95°C for 1 min. 60°C for 1 min., and 72°C for 1 min. (10 cycles); cycle 4 is 95°C for 1 min., 60°C for 1 min., and 72°C for 2 min. (8 cycles); cycle 5 is 72°C for 15 min. (1 cycle); and the final cycle is a 4°C hold until sample is taken out of the machine. The 50-μl PCR products are concentrated down to 10 μl with vacuum centrifugation, and a

sample is then run on a thin 1.2 % Tris-borate-EDTA agarose gel containing ethidium bromide. A band of expected size would indicate that this gene is present in the tissue assayed. The amount of RNA in the pellet may be quantified in numerous ways, for example, it may be weighed.

Verification of the nucleotide sequence of the PCR products is done by microsequencing. The PCR product is purified with a Qiagen PCR Product Purification Kit (Qiagen, Chatsworth, CA) as described by the manufacturer. One μg of the PCR product undergoes PCR sequencing by using the Taq DyeDeoxy Terminator Cycle sequencing kit in a Perkin-Elmer 9600 PCR machine as described by Applied Biosystems (Foster, CA). The sequenced product is purified using Centri-Sep columns (Princeton Separations, Adelphia, NJ) as described by the company. This product is then analyzed with an ABI model 373A DNA sequencing system (Applied Biosystems) integrated with a Macintosh IIci computer.

Example 2

Bacterial Expression and Purification of the CSG Proteins and Use For Preparing a Monoclonal Antibody

The DNA sequence encoding a polypeptide of the present invention, ATCC # 97201, which one is initially amplified using PCR oligonucleotide primers corresponding to the 5′ sequences of the processed protein (minus the signal peptide sequence) and the vector sequences 3′ to the gene. Additional nucleotides corresponding to the DNA sequence are added to the 5′ and 3′ sequences respectively. The 5′ oligonucleotide primer may contain, for example, a restriction enzyme site followed by nucleotides of coding sequence starting from the presumed terminal amino acid of the processed protein. The 3′ sequence may, for example, contain complementary sequences to a restriction enzyme site and also be followed by nucleotides of the nucleic acid sequence encoding the protein of interest. The restriction

enzyme sites correspond to the restriction enzyme sites on a bacterial expression vector, for example, pQE-9 (Qiagen, Inc. Chatsworth, CA). pQE-9 encodes antibiotic resistance (Amp^r), a bacterial origin of replication (ori), an IPTG-regulatable promoter operator (P/O), a ribosome binding site (RBS), a 6-His tag and restriction enzyme sites. pQE-9 is then digested with the restriction enzymes corresponding to restriction enzyme sites contained in he primer sequences. The amplified sequences are ligated into pQE-9 and inserted in frame with the sequence encoding for the histidine tag and the RBS. The ligation mixture is then used to transform an E. coli strain, for example, M15/rep 4 (Qiagen) by the procedure described in Sambrook, J. et al., Molecular Cloning: A Laboratory Manual, Cold Spring Laboratory Press, (1989). M15/rep4 contains multiple copies of the plasmid pREP4, which expresses the lacI repressor and also confers kanamycin resistance (Kan^r). Transformants are identified by their ability to grow on LB plates and ampicillin/kanamycin resistant colonies are selected. Plasmid DNA is isolated and confirmed by restriction analysis. Clones containing the desired constructs are grown overnight (O/N) in liquid culture in LB media supplemented with both Amp (100 ug/ml) and Kan (25 ug/ml). The O/N culture is used to inoculate a large culture at a ratio of 1:100 to 1:250. The cells are grown to an optical density 600 ($O.D._{600}$) of between 0.4 and 0.6. IPTG ("Isopropyl-B-D-thiogalacto pyranoside") is then added to a final concentration of 1 mM. IPTG induces by inactivating the lacI repressor, clearing the P/O leading to increased gene expression. Cells are grown an extra 3 to 4 hours. Cells are then harvested by centrifugation. The cell pellet is solubilized in the chaotropic agent 6 Molar Guanidine HCl. After clarification, solubilized protein is purified from this solution by chromatography on a Nickel-Chelate column under conditions that allow for tight binding by proteins containing the 6-His tag (Hochuli, E. et al., J.

Chromatography 411:177-184 (1984)). The protein is eluted from the column in 6 molar guanidine HCl pH 5.0 and for the purpose of renaturation adjusted to 3 molar guanidine HCl, 100mM sodium phosphate, 10 mmolar glutathione (reduced) and 2 mmolar glutathione (oxidized). After incubation in this solution for 12 hours the protein is dialyzed to 10 mmolar sodium phosphate.

The protein purified in this manner may be used as an epitope to raise monoclonal antibodies specific to such protein. The monoclonal antibodies generated against the polypeptide the isolated protein can be obtained by direct injection of the polypeptides into an animal or by administering the polypeptides to an animal. The antibodies so obtained will then bind to the protein itself. Such antibodies can then be used to isolate the protein from tissue expressing that polypeptide by the use of an, for example, ELISA assay.

Example 3

Preparation of cDNA Libraries from Colon Tissue

Total cellular RNA is prepared from tissues by the guanidinium-phenol method as previously described (P. Chomczynski and N. Sacchi, Anal. Biochem., 162: 156-159 (1987)) using RNAzol (Cinna-Biotecx). An additional ethanol precipitation of the RNA is included. Poly A mRNA is isolated from the total RNA using oligo dT-coated latex beads (Qiagen). Two rounds of poly A selection are performed to ensure better separation from non-polyadenylated material when sufficient quantities of total RNA are available.

The mRNA selected on the oligo dT is used for the synthesis of cDNA by a modification of the method of Gobbler and Hoffman (Gobbler, U. and B.J. Hoffman, 1983, Gene, 25:263). The first strand synthesis is performed using either Moloney murine sarcoma virus reverse transcriptase (Stratagene) or Superscript II (RNase H minus Moloney murine

reverse transcriptase, Gibco-BRL). First strand synthesis is primed using a primer/linker containing an Xho I restriction site. The nucleotide mix used in the synthesis contains methylated dCTP to prevent restriction within the cDNA sequence. For second-strand synthesis E. coli polymerase Klenow fragment is used and [^{32}P]-dATP is incorporated as a tracer of nucleotide incorporation.

Following 2nd strand synthesis, the cDNA is made blunt ended using either T4 DNA polymerase or Klenow fragment. Eco RI adapters are added to the cDNA and the cDNA is restricted with Xho I. The cDNA is size fractionated over a Sephacryl S-500 column (Pharmacia) to remove excess linkers and cDNAs under approximately 500 base pairs.

The cDNA is cloned unidirectionally into the Eco RI-Xho I sites of either pBluescript II phagemid or lambda Uni-zap XR (Stratagene). In the case of cloning into pBluescript II, the plasmids are electroporated into E.coli SURE competent cells (Stratagene). When the cDNA is cloned into Uni-Zap XR it is packaged using the Gigipack II packaging extract (Stratagene). The packaged phage is used to infect SURE cells and amplified. The pBluescript phagemid containing the cDNA inserts are excised from the lambda Zap phage using the helper phage ExAssist (Stratagene). The rescued phagemid is plated on SOLR E.coli cells (Stratagene).

Preparation of Sequencing Templates

Template DNA for sequencing is prepared by 1) a boiling method or 2) PCR amplification.

The boiling method is a modification of the method of Holmes and Quigley (Holmes, D.S. and M. Quigley, 1981, Anal. Biochem., 114:193). Colonies from either cDNA cloned into Bluescript II or rescued Bluescript phagemid are grown in an enriched bacterial media overnight. 400 µl of cells are centrifuged and resuspended in STET (0.1M NaCl, 10mM TRIS Ph 8.0, 1.0 mM EDTA and 5% Triton X-100) including lysozyme (80 µg/ml) and RNase A (4 µg/ml). Cells are boiled for 40

seconds and centrifuged for 10 minutes. The supernatant is removed and the DNA is precipitated with PEG/NaCl and washed with 70% ethanol (2x). Templates are resuspended in water at approximately 250 ng/μl.

Preparation of templates by PCR is a modification of the method of Rosenthal et al. (Rosenthal, et al., Nucleic Acids Res., 1993, 21:173-174). Colonies containing cDNA cloned into pBluescript II or rescued pBluescript phagemid are grown overnight in LB containing ampicillin in a 96 well tissue culture plate. Two μl of the cultures are used as template in a PCR reaction (Saiki, RK, et al., Science, 239:487-493, 1988; and Saiki, RK, et al., Science, 230:1350-1354, 1985) using a tricine buffer system (Ponce and Micol., Nucleic Acids Res., 1992, 20:1992.) and 200 μM dNTPs. The primer set chosen for amplification of the templates is outside of primer sites chosen for sequencing of the templates. The primers used are 5'-ATGCTTCCGGCTCGTATG-3' which is 5' of the M13 reverse sequence in pBluescript and 5'-GGGTTTTCCCAGTCACGAC-3', which is 3' of the M13 forward primer in pBluescript. Any primers which correspond to the sequence flanking the M13 forward and reverse sequences can be used. Perkin-Elmer 9600 thermocyclers are used for amplification of the templates with the following cycler conditions: 5 min at 94°C (1 cycle); (20 sec at 94°C); 20 sec at 55°C (1 min at 72°C) (30 cycles); 7 min at 72°C (1 cycle). Following amplification the PCR templates are precipitated using PEG/NaCl and washed three times with 70% ethanol. The templates are resuspended in water.

Example 4

Isolation of a Selected Clone From Colon Tissue

Two approaches are used to isolate a particular clone from a cDNA library prepared from human colon tissue.

In the first, a clone is isolated directly by screening the library using an oligonucleotide probe. To isolate a

particular clone, a specific oligonucleotide with 30-40 nucleotides is synthesized using an Applied Biosystems DNA synthesizer according to one of the partial sequences described in this application. The oligonucleotide is labeled with ^{32}P- -ATP using T4 polynucleotide kinase and purified according to the standard protocol (Maniatis et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press, Cold Spring, NY, 1982). The Lambda cDNA library is plated on 1.5% agar plate to a density of 20,000-50,000 pfu/150 mm plate. These plates are screened using Nylon membranes according to the standard phage screening protocol (Stratagene, 1993). Specifically, the Nylon membrane with denatured and fixed phage DNA is prehybridized in 6 x SSC, 20 mM NaH_2PO_4, 0.4% SDS, 5 x Denhardt's 500 μg/ml denatured, sonicated salmon sperm DNA; and 6 x SSC, 0.1% SDS. After one hour of prehybridization, the membrane is hybridized with hybridization buffer 6 x SSC, 20 mM NaH_2PO_4, 0.4% SDS, 500 μg/ml denatured, sonicated salmon sperm DNA with 1 x 10^6 cpm/ml ^{32}P-probe overnight at 42°C. The membrane is washed at 45-50°C with washing buffer 6 x SSC, 0.1% SDS for 20-30 minutes dried and exposed to Kodak X-ray film overnight. Positive clones are isolated and purified by secondary and tertiary screening. The purified clone sequenced to verify its identity to the partial sequence described in this application.

An alternative approach to screen the cDNA library prepared from human colon tissue is to prepare a DNA probe corresponding to the entire partial sequence. To prepare a probe, two oligonucleotide primers of 17-20 nucleotides derived from both ends of the partial sequence reported are synthesized and purified. These two oligonucleotides are used to amplify the probe using the cDNA library template. The DNA template is prepared from the phage lysate of the cDNA library according to the standard phage DNA preparation protocol (Maniatis et al.). The polymerase chain reaction is

carried out in 25 μl reaction mixture with 0.5 μg of the above cDNA template. The reaction mixture is 1.5-5 mM $MgCl_2$, 0.01% (w/v) gelatin, 20 μM each of dATP, dCTP, dGTP, dTTP, 25 pmol of each primer and 0.25 Unit of Taq polymerase. Thirty five cycles of PCR (denaturation at 94°C for 1 min; annealing at 55°C for 1 min; elongation at 72°C for 1 min) are performed with the Perkin-Elmer Cetus automated thermal cycler. The amplified product is analyzed by agarose gel electrophoresis and the DNA band with expected molecular weight is excised and purified. The PCR product is verified to be the probe by subcloning and sequencing the DNA product. The probe is labeled with the Multiprime DNA Labelling System (Amersham) at a specific activity < 1 x 10^9 dmp/μg. This probe is used to screen the lambda cDNA library according to Stratagene's protocol. Hybridization is carried out with 5X TEN 920XTEN:0.3M Tris-HCl pH 8.0, 0.02M EDTA and 3MNaCl), 5X Denhardt's, 0.5% sodium pyrophosphate, 0.1% SDS, 0.2 mg/ml heat denatured salmon sperm DNA and 1 x 10^6 cpm/ml of [^{32}P]-labeled probe at 55°C for 12 hours. The filters are washed in 0.5X TEN at room temperature for 20-30 min., then at 55°C for 15 min. The filters are dried and autoradiographed at -70°C using Kodak XAR-5 film. The positive clones are purified by secondary and tertiary screening. The sequence of the isolated clone are verified by DNA sequencing.

General procedures for obtaining complete sequences from partial sequences described herein are summarized as follows;

Procedure 1

Selected human DNA from the partial sequence clone (the cDNA clone that was sequenced to give the partial sequence) is purified e.g., by endonuclease digestion using Eco-R1, gel electrophoresis, and isolation of the clone by removal from low melting agarose gel. The isolated insert DNA, is radiolabeled e.g., with ^{32}P labels, preferably by nick translation or random primer labeling. The labeled insert is used as a probe to screen a lambda phage cDNA library or a

plasmid cDNA library. Colonies containing clones related to the probe cDNA are identified and purified by known purification methods. The ends of the newly purified clones are nucleotide sequenced to identify full length sequences. Complete sequencing of full length clones is then performed by Exonuclease III digestion or primer walking. Northern blots of the mRNA from various tissues using at least part of the deposited clone from which the partial sequence is obtained as a probe can optionally be performed to check the size of the mRNA against that of the purported full length cDNA.

The following procedures 2 and 3 can be used to obtain full length genes or full length coding portions of genes where a clone isolated from the deposited clone mixture does not contain a full length sequence. A library derived from human colon tissue or from the deposited clone mixture is also applicable to obtaining full length sequences from clones obtained from sources other than the deposited mixture by use of the partial sequences of the present invention.

Procedure 2

RACE Protocol For Recovery of Full-Length Genes

Partial cDNA clones can be made full-length by utilizing the rapid amplification of cDNA ends (RACE) procedure described in Frohman, M.A., Dush, M.K. and Martin, G.R. (1988) Proc. Nat'l. Acad. Sci. USA, 85:8998-9002. A cDNA clone missing either the 5' or 3' end can be reconstructed to include the absent base pairs extending to the translational start or stop codon, respectively. In most cases, cDNAs are missing the start of translation therefor. The following briefly describes a modification of this original 5' RACE procedure. Poly A+ or total RNA is reverse transcribed with Superscript II (Gibco/BRL) and an antisense or complementary primer specific to the cDNA sequence. The primer is removed from the reaction with a Microcon Concentrator (Amicon). The

first-strand cDNA is then tailed with dATP and terminal deoxynucleotide transferase (Gibco/BRL). Thus, an anchor sequence is produced which is needed for PCR amplification. The second strand is synthesized from the dA-tail in PCR buffer, Taq DNA polymerase (Perkin-Elmer Cetus), an oligo-dT primer containing three adjacent restriction sites (XhoI, SalI and ClaI) at the 5' end and a primer containing just these restriction sites. This double-stranded cDNA is PCR amplified for 40 cycles with the same primers as well as a nested cDNA-specific antisense primer. The PCR products are size-separated on an ethidium bromide-agarose gel and the region of gel containing cDNA products the predicted size of missing protein-coding DNA is removed. cDNA is purified from the agarose with the Magic PCR Prep kit (Promega), restriction digested with XhoI or SalI, and ligated to a plasmid such as pBluescript SKII (Stratagene) at ShoI and EcoRV sites. This DNA is transformed into bacteria and the plasmid clones sequenced to identify the correct protein-coding inserts. Correct 5' ends are confirmed by comparing this sequence with the putatively identified homologue and overlap with the partial cDNA clone.

Several quality-controlled kits are available for purchase. Similar reagents and methods to those above are supplied in kit form from Gibco/BRL. A second kit is available from Clontech which is a modification of a related technique, SLIC (single-stranded ligation to single-stranded cDNA) developed by Dumas et al. (Dumas, J.B., Edwards, M., Delort, J. and Mallet, Jr., 1991, Nucleic Acids Res., 19:5227-5232). The major differences in procedure are that the RNA is alkaline hydrolyzed after reverse transcription and RNA ligase is used to join a restriction site-containing anchor primer to the first-strand cDNA. This obviates the necessity for the dA-tailing reaction which results in a polyT stretch that is difficult to sequence past.

An alternative to generating 5′ cDNA from RNA is to use cDNA library double-stranded DNA. An asymmetric PCR-amplified antisense cDNA strand is synthesized with an antisense cDNA-specific primer and a plasmid-anchored primer. These primers are removed and a symmetric PCR reaction is performed with a nested cDNA-specific antisense primer and the plasmid-anchored primer.

Procedure 3

RNA Ligase Protocol For Generating The 5′ End Sequences To Obtain Full Length Genes

Once a gene of interest is identified, several methods are available for the identification of the 5′ or 3′ portions of the gene which may not be present in the original deposited clone. These methods include but are not limited to filter probing, clone enrichment using specific probes and protocols similar and identical to 5′ and 3′ RACE. While the full length gene may be present in a library and can be identified by probing, a useful method for generating the 5′ end is to use the existing sequence information from the original partial sequence to generate the missing information. A method similar to 5′ RACE is available for generating the missing 5′ end of a desired full-length gene. (This method was published by Fromont-Racine et al, Nucleic Acids Res., 21(7):1683-1684 (1993). Briefly, a specific RNA oligonucleotide is ligated to the 5′ ends of a population of RNA presumably containing full-length gene RNA transcript and a primer set containing a primer specific to the ligated RNA oligonucleotide. A primer specific to a known sequence (EST) of the gene of interest is used to PCR amplify the 5′ portion of the desired full length gene which may then be sequenced and used to generate the full length gene. This method starts with total RNA isolated from the desired source, poly A RNA may be used but is not a prerequisite for this procedure. The RNA preparation may then be treated with

phosphatase if necessary to eliminate 5′ phosphate groups on degraded or damaged RNA which may interfere with the later RNA ligase step. The phosphatase if used is then inactivated and the RNA is treated with tobacco acid pyrophosphatase in order to remove the cap structure present at the 5′ ends of messenger RNAs. This reaction leaves a 5′ phosphate group at the 5′ end of the cap-cleaved RNA which can then be ligated to an RNA oligonucleotide using T4 RNA ligase. This modified RNA preparation can then be used as a template for first strand cDNA synthesis using a gene-specific oligonucleotide. The first stand synthesis reaction can then be used as a template for PCR amplification of the desired 5′ end using a primer specific to the ligated RNA oligonucleotide and a primer specific to the known sequence (EST) of the gene of interest. The resultant product is then sequenced and analyzed to confirm that the 5′ end sequence belongs to the partial sequence.

Example 5

Expression via Gene Therapy

Fibroblasts are obtained from a subject by skin biopsy. The resulting tissue is placed in tissue-culture medium and separated into small pieces. Small chunks of the tissue are placed on a wet surface of a tissue culture flask, approximately ten pieces in each flask. The flask is turned upside down, closed tight and left at room temperature over night. After 24 hours at room temperature, the flask is inverted and the chunks of tissue remain fixed to the bottom of the flask and fresh media (e.g., Ham's F12 media, with 10% FBS, penicillin and streptomycin, is added. This is then incubated at 37°C for approximately one week. At this time, fresh media is added and subsequently changed every several days. After an additional two weeks in culture, a monolayer of fibroblasts emerges. The monolayer is trypsinized and scaled into larger flasks.

pMV-7 (Kirschmeier, P.T. et al, DNA, 7:219-25 (1988) flanked by the long terminal repeats of the Moloney murine sarcoma virus, is digested with EcoRI and HindIII and subsequently treated with calf intestinal phosphatase. The linear vector is fractionated on agarose gel and purified, using glass beads.

The cDNA encoding a polypeptide of the present invention is amplified using PCR primers which correspond to the 5' and 3' end sequences respectively. The 5' primer contains an EcoRI site and the 3' primer contains a HindIII site. Equal quantities of the Moloney murine sarcoma virus linear backbone and the EcoRI and HindIII fragment are added together, in the presence of T4 DNA ligase. The resulting mixture is maintained under conditions appropriate for ligation of the two fragments. The ligation mixture is used to transform bacteria HB101, which are then plated onto agar-containing kanamycin for the purpose of confirming that the vector had the gene of interest properly inserted.

The amphotropic pA317 or GP+am12 packaging cells are grown in tissue culture to confluent density in Dulbecco's Modified Eagle's Medium (DMEM) with 10% calf serum (CS), penicillin and streptomycin. The mMSV vector containing the gene is then added to the media and the packaging cells are transduced with the vector. The packaging cells now produce infectious viral particles containing the gene (the packaging cells are now referred to as producer cells).

Fresh media is added to the transduced producer cells, and subsequently, the media is harvested from a 10 cm plate of confluent producer cells. The spent media, containing the infectious viral particles, is filtered through a millipore filter to remove detached producer cells and this media is then used to infect fibroblast cells. Media is removed from a sub-confluent plate of fibroblasts and quickly replaced with the media from the producer cells. This media is removed and replaced with fresh media. If the titer of virus

is high, then virtually all fibroblasts will be infected and no selection is required. If the titer is very low, then it is necessary to use a retroviral vector that has a selectable marker, such as neo or his.

The engineered fibroblasts are then injected into the host, either alone or after having been grown to confluence on cytodex 3 microcarrier beads. The fibroblasts now produce the protein product.

Numerous modifications and variations of the present invention are possible in light of the above teachings and, therefore, within the scope of the appended claims, the invention may be practiced otherwise than as particularly described.